Computer Lab 2
Statistics for Business (STAT:1030, Bognar)

Homework: Inference for μ (σ unknown) using R

We will analyze the body temperatures of 130 randomly selected people. Assume that body temperatures, X, can be modeled by a normal distribution with mean μ and standard deviation σ, i.e. $X \sim N(\mu, \sigma)$. We do not know σ. You will do the analysis in R: specifically, you will make a histogram of the data, find a 99% CI for μ, and test $H_0 : \mu = 98.6$ versus $H_a : \mu \neq 98.6$ at the $\alpha = 0.01$ significance level. Print your R commands and output (see below), print your histogram, and answer the following:

1. Based upon the histogram, is it plausible (roughly speaking) that the data came from a normal distribution (population)?
2. On the output, clearly mark the sample size n, the sample mean $\bar{x}$, the sample standard deviation s, and the estimated standard error of the mean $\widehat{se}(\bar{x}) = s/\sqrt{n}$.
3. On the output, clearly mark the 99% CI for μ. Is μ significantly different from 98.6? Why?
4. On the output, clearly mark the test statistic t^* and $p-$value for the hypothesis test. Do you reject H_0? Why?
5. By hand, find a 99% CI for μ (because the degrees of freedom is greater than 100, replace $t_{\alpha/2,n-1}$ with $z_{\alpha/2}$). Compare your results to the R output. *Show your work using clear notation.*
6. By hand, test $H_0 : \mu = 98.6$ versus $H_a : \mu \neq 98.6$ at the $\alpha = 0.01$ significance level. Compute the test statistic and critical value (because the degrees of freedom is greater than 100, replace $t_{\alpha/2,n-1}$ with $z_{\alpha/2}$), plot the rejection region, and state your decision and final conclusion. Compare your results to the R output. *Show your work using clear notation.*
7. Is the normal assumption needed to do inference for μ? Why?

R Commands

Input the data

Open R (see `R-lab1.pdf` on our website if you need to refresh your memory). To load the data into an object called `temp`, we use vector notation, i.e. `c(my data separated by commas)`. The `c` character stands for *combine*. The assignment operator is a left arrow `<-` (i.e. a less than sign followed by a dash). The full command is

```
temp <- c(96.3, 96.7, 96.9, 97.0, 97.1, 97.1, 97.1, 97.2, 97.3, 97.4,
97.4, 97.4, 97.4, 97.5, 97.5, 97.6, 97.6, 97.6, 97.7, 97.8, 97.8,
97.8, 97.8, 97.9, 97.9, 98.0, 98.0, 98.0, 98.0, 98.0, 98.0, 98.1,
98.1, 98.2, 98.2, 98.2, 98.2, 98.3, 98.3, 98.4, 98.4, 98.4, 98.4,
98.5, 98.5, 98.6, 98.6, 98.6, 98.6, 98.6, 98.6, 98.7, 98.7, 98.8,
98.8, 98.8, 98.9, 99.0, 99.0, 99.0, 99.1, 99.2, 99.3, 99.4, 99.5,
96.4, 96.7, 96.8, 97.2, 97.2, 97.4, 97.6, 97.7, 97.7, 97.8, 97.8,
97.8, 97.9, 97.9, 97.9, 98.0, 98.0, 98.0, 98.0, 98.0, 98.1, 98.2,
98.2, 98.2, 98.2, 98.2, 98.2, 98.3, 98.3, 98.3, 98.4, 98.4, 98.4,
98.4, 98.4, 98.5, 98.6, 98.6, 98.6, 98.6, 98.7, 98.7, 98.7, 98.7,
98.7, 98.7, 98.8, 98.8, 98.8, 98.8, 98.8, 98.8, 98.8, 98.9, 99.0,
99.0, 99.1, 99.1, 99.2, 99.2, 99.3, 99.4, 99.9, 100.0, 100.4)
```

You can see the data inside of `temp` by typing its name.

```
temp
```

Remember — you can recall and edit previous commands by using the 'up arrow' on your keyboard

Plot the data

Make a histogram of the dataset:

```
hist(temp)
```

Summary statistics

Find n, $\bar{x}$, s, and $\widehat{se}(\bar{x}) = \frac{s}{\sqrt{n}}$:

```
length(temp)
mean(temp)
sd(temp)
se <- sd(temp)/sqrt(length(temp))
```

Print out the value of $\widehat{se}(\bar{x})$ stored in `se`:

```
se
```

Execute hypothesis test and CI

Test $H_0 : \mu = 98.6$ vs $H_a : \mu \neq 98.6$ at the $\alpha = 0.01$ significance level (and obtain a 99% CI for μ):

```
t.test(temp, alternative=c("two.sided"), mu=98.6, conf.level=0.99)
```

Quit *R*

To quit *R*:

```
q()
```